



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/54, C12N 9/10, C12P 19/26	A2	(11) International Publication Number: WO 00/11190 (43) International Publication Date: 02 March 2000 (02.03.2000)
(21) International Application Number: PCT/CH99/00365 (22) International Filing Date: 06 August 1999 (06.08.1999) (30) Priority Data: 1717/98 20 August 1998 (20.08.1998) CH (60) Parent Application or Grant BERGER, Eric, G. [/]; (). HENNET, Thierry [/]; (). BERGER, Eric, G. [/]; (). HENNET, Thierry [/]; (). BERGER, Eric, G. ; ().		Published
(54) Title: UDP-N-ACETYL GLUCOSAMINYL: 'beta'1,4 GALACTOSIDE: 'beta'1,3N-ACETYL GLUCOSAMINYL TRANSFERASE (POLYLACTOSAMINYL TYPE) (54) Titre: UDP-N-ACETYL GLUCOSAMINYL: 'beta'1,4 GALACTOSIDE: 'beta'1,3N-ACETYL GLUCOSAMINYL TRANSFERASE (TYPE POLYLACTOSAMINYL) (57) Abstract <p>The invention relates to a new gene which codes for UDP-N-acetyl glucosaminy: 'beta'1,4 galactoside: 'beta'1,3(N-acetyl glucosaminy transferase (3GnT). Its enzymatic activity and specificity differ from those of all glycosyl transferases published so far. The invention further relates to isolated DNA molecules and constructs, including their derivatives, which code for 3GnT (referred to as 3GnT-DNA), cloning and expression vectors containing the above 3GnT-DNA, cells transfected with said vectors and recombination methods for obtaining said DNA. The invention also relates to a method for the synthesis of di-, oligo- and polysaccharides with the repetitive element N-acetyl glucosamine ('beta'1}3)galactose('beta'1}4) by obtaining and using the recombinant enzyme from transfected cells which were transfected with 3GnT-DNA, including its derivatives. The invention also relates to methods for diagnosing genetic changes in animals and humans and to the production of antisense samples and targeting vectors for producing transgenic animals.</p> (57) Abrégé <p>L'invention concerne un nouveau gène qui code pour la UDP-N-acétylglucosaminy: 'beta'1,4 galactoside: 'beta'1,3N-acétylglucosaminytransférase (3GnT) et dont l'activité enzymatique et la spécificité diffèrent de celles de toutes les glucosyltransférases publiées à ce jour. L'invention concerne des molécules et structures d'ADN isolées, y compris leurs dérivés, codant pour la 3GnT (désignée comme 3GnT-ADN); des vecteurs de clonage et d'expression contenant cette 3GnT-ADN ainsi que des cellules transfectées par ces vecteurs et des procédés de recombinaison permettant d'obtenir cet ADN. L'invention concerne également un procédé de synthèse de di-, oligo-, polysaccharides ayant l'élément répétitif N-acétylglucosamine ('beta'1}3) Galactose ('beta'1}4) en obtenant et utilisant l'enzyme recombinant à partir de cellules transfectées par la 3GnT-ADN, y compris ses dérivés. L'invention concerne également un procédé permettant de diagnostiquer des modifications génétiques chez l'homme et les animaux, la production d'échantillons antisens et de vecteurs de _ciblage _ permettant de produire des animaux transgéniques.</p>		

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro



INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁷ : C12N 15/54, 9/10, C12P 19/26		A2	(11) Internationale Veröffentlichungsnummer: WO 00/11190
			(43) Internationales Veröffentlichungsdatum: 2. März 2000 (02.03.00)
(21) Internationales Aktenzeichen: PCT/CH99/00365		(81) Bestimmungsstaaten: CA, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) Internationales Anmeldedatum: 6. August 1999 (06.08.99)			
(30) Prioritätsdaten: 1717/98 20. August 1998 (20.08.98) CH		Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>	
(71)(72) Anmelder und Erfinder: BERGER, Eric, G. [CH/CH]; Universität Zürich, Physiologisches Institut, Winterthurerstrasse 190, CH-8057 Zürich (CH). HENNET, Thierry [CH/CH]; Universität Zürich, Physiologisches Institut, Winterthurerstrasse 190, CH-8057 Zürich (CH).			
(74) Gemeinsamer Vertreter: BERGER, Eric, G.; Universität Zürich, Physiologisches Institut, Winterthurerstrasse 190, CH-8057 Zürich (CH).			
(54) Title: UDP-N-ACETYL GLUCOSAMINYL: β 1,4 GALACTOSIDE: β 1,3N-ACETYL GLUCOSAMINYL TRANSFERASE (POLYLACTOSAMINYL TYPE)			
(54) Bezeichnung: UDP-N-AZETYL GLUKOSAMINYL: β 1,4 GALAKTOSID: β 1,3N-AZETYL GLUKOSAMINYL TRANSFERASE (POLYLAKTOSAMINYL TYP)			
(57) Abstract			
<p>The invention relates to a new gene which codes for UDP-N-acetyl glucosaminyl: β1,4 galactoside: β1,3(N-acetyl glucosaminyl transferase (3GnT). Its enzymatic activity and specificity differ from those of all glycosyl transferases published so far. The invention further relates to isolated DNA molecules and constructs, including their derivatives, which code for 3GnT (referred to as 3GnT-DNA), cloning and expression vectors containing the above 3GnT-DNA, cells transfected with said vectors and recombination methods for obtaining said DNA. The invention also relates to a method for the synthesis of di-, oligo- and polysaccharides with the repetitive element N-acetyl glucosamine (β1\rightarrow3)galactose(β1\rightarrow4) by obtaining and using the recombinant enzyme from transfected cells which were transfected with 3GnT-DNA, including its derivatives. The invention also relates to methods for diagnosing genetic changes in animals and humans and to the production of antisense samples and targeting vectors for producing transgenic animals.</p>			
(57) Zusammenfassung			
<p>Gegenstand der Erfindung ist ein neues Gen, welches für die UDP-N-Acetylglukosaminyl: β1,4 Galaktosid: β1,3N-Acetylglukosaminyltransferase (3GnT) kodiert. Die enzymatische Aktivität und Spezifität ist von allen bisher publizierten Glykosyltransferasen verschieden. Die Erfindung umfasst isolierte DNA-Moleküle und -Konstrukte einschliesslich ihrer Derivate, welche für die 3GnT kodieren (als 3GnT-DNA bezeichnet), wie auch Klonierungs- und Expressionsvektoren, welche diese 3GnT-DNA enthalten, sowie Zellen, welche mit diesen Vektoren transfektiert sind und Methoden der Rekombination zur Gewinnung dieser DNA. Die Erfindung betrifft zudem Verfahren zur Synthese von Di-, Oligo- und Polysacchariden mit dem repetitiven Element N-acetylglukosamin(β1\rightarrow3)Galaktose(β1\rightarrow4) durch Gewinnung und Anwendung des rekombinanten Enzyms aus transfektierten Zellen, die mit 3GnT-DNA, einschliesslich ihrer Derivate, transfektiert sind. Die Erfindung umfasst auch Verfahren zur Diagnose genetischer Änderungen bei Tier und Mensch, der Herstellung von Antisense-Proben und "Targeting" Vektoren zur Erzeugung genveränderter Tiere.</p>			

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Description

5

10

15

20

25

30

35

40

45

50

55

UDP-N-Azetvlglukosaminvl:β1,4 Galaktosid: β1,3N-Azetvlglukosaminvltrans-
ferase (Polvlaktosaminvl Typ)

Technisches Gebiet

Die vorliegende Erfindung betrifft das Gebiet C (Chemie, Hüttenwesen). Im engeren Sinn betrifft die Erfindung das Gebiet der Biotechnologie und der rekombinanten DNA-Technik. Die Erfindung betrifft die Klonierung einer Nukleinsäure, welche für eine neuartige Glykosyltransferase kodiert. Die Glykosyltransferasen zählen zu derjenigen Klasse von Enzymen, welche an den Glykosylierungsmechanismen beteiligt sind. Diese gehören zu den wichtigsten biosynthetischen Eigenschaften der lebenden Zelle und sind vor allem an posttranslatorischen Veränderungen von Proteinen beteiligt. Zudem berühren die erfindungsgemässen Nukleinsäuren auch die Synthese der Glykosaminoglykane (oder, synonym dazu, Proteoglykane).

Im Besonderen betrifft die Erfindung eine Familie von ähnlichen DNA¹-Sequenzen (Nukleinsäuren), von welchen eine die Spezifität aufweist, welche sie zusammen mit einer β1,4Galaktosyltransferase (Almeida, R. et al., J. Biol. Chem. 1997, 272, 31979-91; Sato T., et al., Proc. Natl. Acad. Sci. U S A 1998, 95, 472-7; Sato T. et al, Biochem. Biophys Res. Commun. 1998, 244, 637-641) zur Synthese von Polyaktosaminoglykanen bzw. Keratan befähigt. Die Erfindung bezieht sich auf das Gen, die daraus in vitro abgeleitete cDNA, DNA Konstrukte mit ähnlicher Aktivität, rekombinante Plasmide, welche dieses Gen, bzw. cDNA enthalten, sowie Produktionsplasmide zur Herstellung des rekombinanten Enzyms mit Hilfe der Transfektion und Transformation sowie alle Methoden zur Identifikation von Polymorphismen bei Menschen und Tieren und der Herstellung von Antisense-Polynukleotiden und sogenannten Targeting Vektoren zur Herstellung von genveränderten Tieren.

¹ DNA: In diesem Schriftstück steht DNA für Deoxyribonucleic Acid, zu deutsch Desoxyribonukleinsäure

5

2

Stand der Technik

10

Alle Literaturzitate in diesem Schriftstück beziehen sich auf die Gesamtheit der zitierten Literaturstelle.

5

15

Die Glykosylierung ist eine der wichtigsten posttranslatorischen Veränderungen der kovalenten Proteinstruktur überhaupt. Sie beginnt nach der eigentlichen ribosomalen Proteinsynthese und der Translokation des neugebildeten Proteins in das endoplasmatische Reticulum durch die Uebertragung eines präformierten Oligosaccharids ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$)² auf das Protein. Dieser Mechanismus, der in der Evolution von der Bäckerhefe bis zum menschlichen Organismus sehr ähnlich abläuft und unter dem Begriff der *N*-Glykosylierung in der Literatur eingehend beschrieben ist (Glycoproteins, Montreuil J., Vliegthart, J.F.G. & Schachter, H., Hsg. Elsevier, Lausanne, 1995), bildet die Vorstufe zum Glykosylierungsprozess, welcher im Golgi-Apparat stattfindet. Im Golgi Apparat findet zudem noch die sog. *O*-Glykosylierung statt. Die Glykosylierungsprozesse im Golgi-Apparat führt zur Ausbildung spezifischer Glykanstrukturen, welche bei zahlreichen biochemischen und physiologischen Phänomenen beteiligt sind (Varki, A., Glycobiology 1993, 3, 97-130). Sowohl *O*- wie *N*-gebundene Glykane weisen in gewissen Zellen und bei bestimmten Entwicklungsstufen verlängerte Antennen auf, welche aus der Repetition eines Disaccharid-Motivs entstehen, nämlich des $\text{Gal}\beta 1,4\text{GlcNAc}\beta 1,3\text{Gal}\beta 1,4\text{GlcNAc}^3$ (Glycoproteins and Disease, Montreuil J., Vliegthart, J.F.G. & Schachter, H., Hsg., Elsevier, Lausanne, 1996). Dieses Motiv führt in variabler Länge zu einer Struktur, welche zusammenfassend als Poly-laktosaminoglykane bezeichnet werden. Die gleiche Struktur findet sich als Glykangrundgerüst im Keratan, einem Glykosaminoglykan mit bestimmter Verteilung im Organ (Cornea, Knorpel) (Rodan L. in The Biochemistry of Glycoproteins and Proteoglycans, Lennarz, W.J., Hsg. Plenum Press, New York, 1980, S 314-318).

25

15

30

20

35

40

45

30

Die Poly-laktosaminoglykane sind auch bei krankhaften Prozessen gefunden worden, wie bei der viralen Transformation von Zellen (Kobata A. in Glycoproteins and Disease, Montreuil J., Vliegthart, J.F.G. & Schachter, H., Hsg., Elsevier, Lausanne, 1996, S 211-227) und bei der Karzinogenese (Yamashita K. & Kobata A., ibidem, S 197-210). Maligne transformierte Zellen erwerben neuartige Adhäsions und Erkennung-Eigenschaften im Verbund mit anderen Zellen (Metasta-

50

² $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ heisst Glukose₃Mannose₉N'-azetylglukosamin₂

³ $\text{Gal}\beta 1,4\text{GlcNAc}\beta 1,3\text{Gal}\beta 1,4\text{GlcNAc}$ heisst Galaktose $\beta 1 \Rightarrow 4$ N'-azetylglukosamin $\beta 1,3$ Galaktose $\beta 1 \Rightarrow 4$ N'-azetylglukosamin

55

sierung und Invasivität), bei welchen das Glykanmuster der malignen Zelle von grosser Bedeutung ist (Yamashita K. & Kobata A., *ibidem*, S. 229-239).

Bei der Entstehung der Entzündung spielen Glykan-Selectin Wechselwirkungen eine entscheidende Rolle: So ist beschrieben worden, dass Leukocyten ein Glykoprotein enthalten (PSGL-1)⁴,

5 welches eine bestimmte O-glykosidische Struktur aufweist, welches Polyaktosaminoglykane enthält (Liu, W. et al., *J. Biol. Chem.* 1998, 273, 7078-7087). Diese sind bei der Erkennung des P-Selectins auf der endothelialen Oberfläche der Gefässe, welche das entzündete Gebiet versorgen, beteiligt (McEver, R. and Cummings, R.D., *J. Clin. Invest.* 1997, 100, 485-492}. Deshalb fokussieren einige pharmazeutische Firmen ihre Anstrengungen auf das Gebiet dieser Wechselwirkung, mit dem Ziel, spezifische Inhibitoren zu entwickeln.

Daraus ergibt sich ein Anwendungsgebiet dieser Enzyme, welche in vivo an der Synthese biospezifischer Erkennungsstrukturen beteiligt sind: Sie können für die in vitro Synthese ähnlicher Strukturen nutzbar gemacht werden.

Ein weiteres Anwendungsgebiet liegt in der Herstellung von Glycosaminoglycanen (Proteoglycane). Diese bilden einen wesentlichen Teil der extrazellulären Matrix. Zu diesen Stoffen zählt Keratan und seine Derivate. Diese enthalten im Glycanteil als Bauelement das Disaccharid N-Azetylglukosamin β 1 \Rightarrow 3Galaktose β 1 \Rightarrow 4. Dieses Disaccharid bildet als repetitives Strukturelement Glykanketten von variabler Länge. Diese konnten bisher in vitro enzymatisch nicht synthetisiert werden, da das Enzym, welches effizient die N-Azetylglukosamin β 1 \Rightarrow 3Galaktose Bindung katalysiert, nicht verfügbar war.

⁴ PSGL-1 steht für P-Selectin Glykoprotein Ligand-1

5

4

Beschreibung der Erfindung

10

Die vorliegende Erfindung beschreibt die Isolierung humaner und muriner DNA, welche für die
5 UDP-*N*-Azetylglukosaminyl:β1,4 Galaktosid: β1,3*N*-Azetylglukosaminyltransferase (Polylaktos-
aminyl Typ, nachfolgend kurz als 3GnT bezeichnet) kodiert mit der zugehörigen cDNA und der
15 genomischen DNA. Die Spezifität des Enzyms bezieht sich auf das Donor-Substrat (UDP-*N*-
Azetylglukosamin) und das Akzeptorsubstrat *N*-Azetyllaktosamin und alle Glykokonjugate, wel-
che diese Struktur terminal exponieren.

20

10 Die komplette Nukleotidsequenz ist auf Abb. 1 angegeben.
Eine erste Ausführungsform der Erfindung betrifft die Nukleotidsequenz der humanen Variante
Nukleotid 1 bis 990 (Seq ID-Nr. 1a), welche als konserviert bezeichnet wird und alle Funktions-
konservierten Sequenz-Varianten. Eine zweite Ausführungsform der Erfindung betrifft die Nu-
kleotidsequenz der murinen Form des Enzyms (Seq ID-Nr. 1b) und alle Funktions-konservierten
25 Sequenz-Varianten. Auch betrifft die Erfindung isolierte Nukleinsäuren, welche mit den Sequen-
zen Nr. 1 hybridisieren können einschliesslich daraus abgeleitete Fragmente oder Varianten mit
konservierter Sequenz oder Funktion. Unter Hybridisierung sind auch Bedingungen eingeschlos-
senen, welche einer mittleren Stringenz entsprechen.

30

Eine dritte Ausführungsform entspricht der Sequenz des humanen Proteins, welches durch
20 Translation¹ der erfindungsgemässen Nukleinsäuresequenz entsprechend der Seq ID-Nr. 1a er-
halten wird. Diese Aminosäuresequenz (Seq ID-Nr. 2a) ist durch das offene Leseraster (ORF,
open reading frame) definiert mit Beginn beim Met₁ (Aminosäure 1) bis Pro₃₂₉ (Aminosäure 329)
35 In einer vierten Ausführungsform entspricht die murine Sequenz (Seq ID Nr. 2b) derjenigen der
Aminosäuren mit Beginn beim Met₁ (Aminosäure 1) bis Pro₃₂₅ (Aminosäure 325), welche durch
25 das offene Leseraster der erfindungsgemässen Nukleinsäuresequenz entsprechend der Seq ID
Nr. 1b definiert ist.

40

In einer fünften Ausführungsform der erfindungsgemässen Sequenz Seq ID Nr. 2a beginnt die
Sequenz bei Arg₃₁ bis Aminosäure Pro₃₂₉ (Seq ID Nr. 3a). In einer sechsten Ausführungsform der
erfindungsgemässen Sequenz Seq ID Nr. 2b beginnt die Sequenz bei Arg₃₁ bis Aminosäure Pro₃₂₅
45 (Seq ID Nr. 3b).

50

¹ Unter Translation wird in diesem Schriftstück eine kollineare Übersetzung der Nukleinsäuresequenz in eine Ami-
nosäuresequenz entsprechend dem universellen genetischen Code bezeichnet.

50

55

5 In einem damit zusammenhängenden Aspekt beschreibt die Erfindung DNA-Konstrukte, welche die 3GnT-Sequenzen enthalten, einschliesslich aber nicht ausschliesslich solcher, welche die 3GnT Sequenz mit einer transkriptions-regulatorischen Sequenz mit oder ohne Polyadenylierungssignal verknüpft. Solche DNA-Konstrukte werden auch als "Vektoren" bezeichnet. Zellen, welche mit
10 diesen Vektoren transient oder stabil transfektiert sind, sind in der Beschreibung eingeschlossen. Ebenso Viren und Bakteriophagen, welche die 3GnT-Sequenzen enthalten, sind in der Beschreibung eingeschlossen. Die Erfindung schliesst auch die Methoden zur Produktion von 3GnT Polypeptiden ein. Diese beruhen a) auf der Einführung des 3GnT Gens, bzw. cDNA oder eines davon abgeleiteten Konstruktes in eine Wirtszelle; b) die für die Gewinnung eines aktiven und/oder immunoreaktiven rekombinanten 3GnT-Polypeptids geeigneten Bedingungen sowie c) für die Isolierung des rekombinanten Produktes. Ebenso schliesst die Erfindung die Herstellung einer Zelllinie ein, welche die 3GnT stabil exprimiert unter Einbezug des dazu erforderlichen Transfers der cDNA entsprechend der Sequenzen gemäss Seq ID Nr. 1a, bzw 1b, oder eines Fragmentes dieser cDNAs, welche für ein Protein entsprechend der Sequenz 2a, bzw 2b codiert oder eines Fragmentes, das enzymatisch aktiv ist oder sich zur Kristallisierung eignet. Eingeschlossen sind die
15 Bedingungen zur Selektion und zum Wachstum der die 3GnT exprimierenden Wirtszellen sowie zur ihrer Identifikation. Diese Zellen können zur Herstellung des rekombinanten Enzyms gebraucht werden, das als Katalysator zur Herstellung von Glykanen verwendet werden kann. Diese cDNA kann auch dazu verwendet werden, Zell-Linien oder ex vivo gewonnene Zellen stabil zu
20 transfektieren und damit ihre Adhäsions- und Erkennungseigenschaften zu verändern. Auch können solchermassen stabil transfektierte Zellen zur Expression von Glykoproteinen eingesetzt werden, welche durch die Expression der 3GnT mit Poly-laktosaminoglykanen substituiert werden. Auch sind alle Anwendungen zur Herstellung von Keratan und seiner Derivate eingeschlossen. Ein weiterer Aspekt der Erfindung betrifft die Isolierung von 3GnT Polypeptiden, welche die Sequenzen Nr. 2a und 2b bzw. 3a und 3b oder funktionskonservierte Teile derselben enthalten oder entsprechende Sequenzen, die mit einem weiteren Peptid im Leseraster fusioniert sind. Unter diesen Peptiden sind ohne Ausschliesslichkeit speziell diejenigen bevorzugt, welche für die Reinigung der rekombinanten 3GnT hilfreich sind.
40 Ein anderer Aspekt der Erfindung betrifft die Methoden, welche zur Entdeckung von Mutationen in der codierenden Region des Gens gebraucht werden, indem genomische DNA von Patienten beispielsweise aus Blutzellen analysiert wird. In einer Ausführungsform umfasst das Verfahren die Isolierung der Patienten-DNA, die Amplifikation mittels PCR von Exonabschnitten des 3GnT-

Gens, die Sequenzierung des amplifizierten Abschnittes und die Feststellung einer möglichen Struktur- oder Promotormutation und die Verknüpfung mit der Pathogenese der Krankheit.

Detaillierte Darstellung der Erfindung

Definitionen:

1. "Nukleinsäuren" oder "Polynukleotide" bezeichnen in diesem Schriftstück Polymere aus Purin- und Pyrimidinbasen einer beliebigen Länge, entweder als Polyribonukleotide, Polydesoxyribonukleotide oder Mischungen beider. Diese umfassen Einzel- und Doppelstränge, beispielsweise DNA-DNA, DNA-RNA und RNA-RNA-Hybride. Ebenso sind atypische Basen mit erwähnt.
2. "cDNA" (Kopie DNA oder Komplementäre DNA) bezeichnet in diesem Schriftstück ein DNA Molekül, eine DNA oder ein davon abgeleiteter Klon, welche enzymatisch aus einer Boten-RNA (mRNA) als Matrize⁶ synthetisiert worden sind. Ein "DNA Konstrukt" ist ein DNA Molekül oder ein Klon eines solchen Moleküls, entweder doppel- oder einzelsträngig, welches Segmente enthält, die in dieser Weise in der Natur nicht vorkommen. Ein nicht begrenzendes Beispiel ist eine cDNA oder eine DNA, welche keine Introns enthält und direkt mit einer exogenen DNA-Sequenz verknüpft ist.
3. Ein "Plasmid" oder generell ein "Vektor" ist ein DNA Konstrukt, welches genetische Information enthält, die zur Replikation in einer Wirtszelle benötigt wird. Solche Plasmide oder Vektoren enthalten im allgemeinen die für die Expression bestimmte Gensequenz und zusätzlich Sequenzen, welche die Expression fördern, unter Einschluss von Promotoren und Transkriptions-Initiationsstellen. Das Plasmid kann linear oder zirkulär angeordnet sein. Im vorliegenden Schriftstück sind als Vektoren auch DNA-Konstrukte eingeschlossen, welche als "targeting vectors" bezeichnet werden. Diese werden zur gezielten Geninaktivierung oder zur Überexpression eines Gens in vivo verwendet.
4. Nukleinsäuren sind miteinander "hybridisierbar", wenn mindestens ein Strang mit einem anderen unter bestimmten Bedingungen der Stringenz einen Doppelstrang bildet. Die Stringenz wird unter anderem bestimmt durch a) die Temperatur während des Hybridisierungsexperimentes mit oder ohne Waschschrift, b) die Ionenstärke und Polarität (zB Formamid) der Hybridisierungs- und Waschlösungen. Die Hybridisierung erfolgt nur bei signifikanter Komplementarität der Basen; je nach Stringenz können allerdings Abweichungen von der Komplementarität toleriert werden. Ty-

⁶ Als Matrize wird in diesem Schriftstück eine RNA- oder DNA-Sequenz bezeichnet, welche für synthetische Reaktionen zur Herstellung von DNA-Kopien verwendet werden; der englische Ausdruck ist "template".

5 pischerweise erfolgt eine Hybridisierung zweier Stränge bei hoher Stringenz (z.B. in wässriger Lösung mit 0,1 x SSC bei 65 °C) nur, wenn eine hohe Komplementarität über die gesamte Länge vorliegt. Mittlere Stringenz (z.B. in wässriger Lösung mit 0,5 x SSC bei 50 °C) und tiefer Stringenz (z.B. in wässriger Lösung mit 1 x SSC bei 40 °C) erfordert entsprechend ein geringeres
10 Mass an Komplementarität zwischen den hybridisierenden Strängen (1x SSC entspricht einer Lösung von 0.15 M NaCl, 0.015 M Na Zitrat).

5. Eine "isolierte" Nukleinsäure oder ein "isoliertes" Polypeptid bezeichnet in diesem Schriftstück eine Substanz der entsprechenden Stoffklasse, welche aus ihrer ursprünglichen Umgebung entfernt worden ist, beispielsweise ihrer natürlichen Umgebung, sofern dieser Stoff in der Natur vor-
15 kommt. Die Isolierung bezeichnet einen Vorgang, bei welchem mindestens 50% der anderen Komponenten aus dem ursprünglichen Gemisch entfernt worden sind, vorzugsweise jedoch über 90%.

6. Als "Sonde" bezeichnet man eine Nukleinsäure, welche mit einem bestimmten Sequenzabschnitt des zu untersuchenden Gens hybridisiert.

7. Eine "abgeleitete" Nukleinsäure bezeichnet eine Nukleinsäuresequenz, welche einem bestimmten Abschnitt einer Sequenz entspricht. Dies sind Sequenzen, welche als homolog, ähnlich, komplementär (d.h. "antisense") oder sequenz-konserviert oder funktions-konserviert bezeichnet sind. Sequenz-konserviert sind solche Varianten, bei welchen ein oder mehrere Änderungen von Nukleotiden eines bestimmten Kodons keine Änderungen der Aminosäuresequenz an dieser Stelle
20 nach sich ziehen. Funktions-konservierte Varianten der 3GnT sind solche, bei welchen Mutationen einzelner Aminosäuren die Konformation und die Enzymaktivität unter Einschluss der Substratspezifität nicht beeinflusst; diese Mutationen betreffen unter anderen Austausch von Aminosäuren mit ähnlichen physiko-chemischen Eigenschaften, wie den sauren, basischen oder hydrophoben oder anderen.

8. Ein "Donor-Substrat" ist ein Molekül, welches von der Transferase erkannt wird und welches für die Enzymkatalyse einen *N*-Azetylglukosamin-Rest beiträgt. Für die 3GnT ist das Donorsubstrat UDP-GlcNAc⁷. Ein "Akzeptorsubstrat" ist ein Molekül, auf welches der GlcNAc-Rest durch die 3GnT übertragen wird. Typischerweise wird dabei der GlcNAc-Rest in einer $\beta 1 \Rightarrow 3$ glykosidischen Bindung auf einen β -anomerisch konfigurierten Galaktose-Rest übertragen. Dieser Galaktose-Rest ist beispielsweise Teil eines *N*- oder *O*-gebundenen Glykans, ohne andere Möglichkeiten auszuschliessen.
45

⁷ UDP-GlcNAc steht für Uridindiphospho-*N*-azetylglukosamin

5 9. Glykokonjugate sind Biomoleküle, welche komplex verknüpfte Glykane enthalten. Diese Stoffe umfassen Glykoproteine, Glykolipide und Proteoglykane (bzw. Glykosaminoglykane).

Die vorliegende Erfindung betrifft ein isoliertes DNA-Molekül unter Einschluss genomischer DNA und der cDNA, welche für eine UDP-*N*-Azetylglucosaminyl:β1,4 Galaktosid: β1,3/*N*-

10 5 Azetylglucosaminyltransferase (Polylaktosaminyl Typ) kodiert. Die cDNA der 3GnT wurde so kloniert, dass die Sequenzähnlichkeiten einer Familie von β1,3Galaktosyltransferasen als Grundlage zur Identifikation von funktions-undefinierten EST⁸-Sequenzen verwendet worden sind. Mit solchen Sequenzen wurde eine in unserem Labor entwickelte menschliche Genbank auf das Vorliegen ähnlicher Sequenzen geprüft. Eine Sequenz, welche ein offenes Leseraster aufwies, wurde
15 in einem Expressionsvektor zusammen mit einer spezifischen Bindungssequenz von 6 Histidinen⁹ in Insekten-Zellen exprimiert. Das exprimierte Produkt, welches den *His-tag* enthielt, wurde in einem Enzymsatz, der als Donorsubstrate sowohl UDP-Galaktose als auch UDP-GlcNAc zusammen mit dem Akzeptorsubstrat LAcNAc¹⁰ enthielt, auf seine Funktion geprüft. Es erwies sich überraschenderweise, dass die gemessene Aktivität ausschliesslich auf die Zugabe des UDP-
20 GlcNAc's zurückzuführen war. Dieser Befund, dass eine cDNA, welche mit einer Genfamilie der β1,3Galaktosyltransferasen (Hennet T. et al., J. Biol. Chem. 1998, 273, 58-65) homolog ist, als Donorsubstrat nicht UDP-Galactose sondern UDP-*N*-Azetylglucosamin erkennt, aber mit den bekannten *N*-Azetylglucosaminyltransferasen keine Homologie aufweist, ist neuartig.

20 *Angaben zu den Reagenzien, DNA, Vektoren und Wirtszellen*

Bei der Erarbeitung vorliegender Erfindung wurden viele herkömmliche Techniken der Molekular- und Zellbiologie verwendet. Diese Techniken entsprechen dem allgemeinen Wissenstand und werden unter anderem im folgenden Standardwerk im Detail beschrieben: Sambrook et al., 1989, Molecular Cloning: A laboratory Manual, 2. Auflage, Cold Spring Harbor Laboratory Press, Cold
25 Spring Harbor, New York.

Die Erfindung umfasst alle Nukleinsäure Sequenzen entsprechend den erfindungsgemässen Sequenzen dargestellt als Seq ID-Nr. 1a, bzw. 1b und davon abgeleitete Fragmente. Die Fragmente sind mindestens 8 Nukleotide lang, bevorzugt sind aber mindestens 12 Nukleotide und noch mehr bevorzugt 15 bis 20 Nukleotide. Die Erfindung umfasst des weiteren alle Nukleinsäurederivate

⁸ EST heisst "Expressed sequence tag" und bezeichnet undefinierte DNA-Sequenzen welche aus der systematischen Erfassung von menschlichen RNA-Molekülen und ihrer reversen Transkription abgeleitet worden sind. Diese EST Sequenzen sind in einer öffentlich zugänglichen Datenbank abrufbar.

⁹ Diese aus 6 linear angeordneten Histidine werden in diesem Schriftstück als *His-tag* bezeichnet

¹⁰ LAcNAc steht für *N*-azetylaktosamin (Galaktoseβ1→4*N*-azetylglucosamin)

entsprechend den erfindungsgemassen Sequenzen, welche unter den Stringenzbedingungen von 1 x SSC, 40 °C mit der Seq ID Nr. 1a bzw 1b hybridisieren, bevorzugt sind Hybridisierungsbedingungen von 0,5 x SSC, 50 °C und noch mehr bevorzugt solche von 0,1 x SSC bei 65 °C.

Die Nukleinsäuren können unmittelbar aus Zellen isoliert werden. Als Alternative kann auch die PCR Methode verwendet werden, um die erfindungsgemassen Nukleinsäuren zu gewinnen, entweder durch Gebrauch von isolierter RT nach reverser Transkription oder genomischer DNA als Matrize. Die für die PCR notwendigen Primer können auf Grund der erfindungsgemassen Sequenzen hergestellt und durch Einführung von Sequenzen für Restriktionsschnittstellen ergänzt werden.

Die erfindungsgemassen Sequenzen können am 5', bzw. 3' Ende natürliche oder heterologe flankierende Sequenzen mit oder ohne Expressions-regulierende Abschnitte wie Promotoren, Enhancers, Bindungsstellen, Polyadenylierungssequenzen, Introne, 3'- oder 5' nicht kodierende Sequenzen und ähnliches enthalten. Die Nukleinsäuren können auch entsprechend dem allgemeinen Wissensstand modifiziert werden wie durch Methylierung, "Capping", Einführen von Nukleinsäure-Analoga, Veränderungen der Phosphodiesterbrücken durch Einführen von ungeladenen Methylphosphonaten, Phosphotriestern, Phosphoramidaten, Carbamate usw. und mit geladenen Brücken (wie z.B. Phosphorothioate, Phosphorodithioate usw). Die Nukleinsäuren können eine oder mehrere kovalente Substituenten besitzen, wie z. B. Proteine (wie Nukleasen, Toxine, Antikörper, Signalpeptide, Polyl-L-Lysine), Interkalatoren (wie z.B. Akridin, Psoralen usw.), Chelatoren (wie Metallchelatoren, radioaktive Metalle, Eisen etc) und alkylierende Substanzen. Die erfindungsgemassen Nukleinsäuren und ihre Derivate können weiter als Methyl- oder Äthyl-phosphotriester oder als Alkylphosphoramidate verändert werden. Ebenso können sie mit einem Marker wie radioaktive Isotope, fluoreszierende Stoffe, Biotin usw. versehen werden.

Brauchbare Sonden, welche auf Grund der erfindungsgemassen Nukleinsäuren entwickelt werden, enthalten also mindestens 8 Nukleotide, welche in der Seq ID 1a bzw. 1b oder Sequenzen- oder Funktions-konservierte Analogsequenzen oder Komplementarsequenz vorkommen und die mit einem oben erwähnten Marker versehen sind.

Die Erfindung betrifft auch Nukleinsäure-Vektoren, welche die erfindungsgemassen Sequenzen oder Derivate und Fragmente davon enthalten. Eine breite Zahl von Vektoren, inklusive Plasmide und Viren sind für die Replikation und/oder Expression in verschiedenen Wirtszellen beschrieben worden; diese können sowohl für die Gentherapie als auch für die Klonierung und für die Proteinexpression eingesetzt werden.

5 Rekombinante Klonierungsvektoren enthalten oft ein oder mehrere Replikationssysteme für die Klonierung oder die Expression, ein oder mehrere Selektionsmarker in der Wirtszelle wie beispielsweise für die Antibiotikaresistenz, dazu eine oder mehrere Expressionskassetten. Die eingefügten Sequenzen können nach Standardmethoden synthetisiert, aus natürlichen Quellen oder als
10 5 Hybride isoliert werden. Die Verknüpfung der kodierenden Abschnitte der erfindungsgemässen Nukleinsäuren mit regulierenden Nukleinsäureelementen oder mit anderen Protein-kodierenden Nukleinsäuren kann mit bekannten Methoden erfolgen. Geeignete Wirtszellen können mit irgendeiner geeigneten Methode transfektiert werden, wie z.B. durch Elektroporation, CaCl_2 -vermittelte
15 DNA Aufnahme, Liposom-verpackte DNA, Infektion mit Pilzen, Mikroinjektion, virale Infektion usw.

Geeignete Wirtszellen umfassen Bakterien, Archaeobakterien, Pilze, vor allem Hefen, Pflanzen und
20 tierische Zellen. Besonders interessant sind *Saccharomyces cerevisiae*, *Saccharomyces pombe*, *Pichia pastoris*, Insekten-Zellen, *Neurospora*, CHO Zellen, COS Zellen, HeLa Zellen und immortalisierte myeloische und lymphoide Zelllinien. Bevorzugte Replikationssysteme umfassen
25 M13, SV40, Baculovirus, Lambda, Retro-und Adenovirus. Eine grosse Zahl von Transkriptionsinitiations- und Terminationssequenzabschnitte sind bekannt und auch nachgewiesenermassen für die Expression von heterologen Proteinen wirksam. Beispiele solcher Abschnitte, ihre Isolierung und Mutagenisierung sind im Fachbereich bekannt. Unter geeigneten Bedingungen können Wirtszellen rekombinante 3GnT und Derivate davon exprimieren und deshalb als Quelle für die Pro-
30 duktion dieses Enzyms verwendet werden.

Mit Vorteil werden DNA-Konstrukte, welchen ein Transkriptions-regulierendes Element zur 3GnT kodierenden Sequenz beigefügt ist, verwendet. Dieser Promotor kann Operatorabschnitte und/oder ribosomale Bindungsstellen enthalten. Nicht erschöpfend aufgezählt fallen darunter
35 bakterielle Promotoren, welche mit *E. coli* kompatibel sind wie β -Laktamase (Penicillinase) Promotor, Laktose Promotor, Tryptophan Promotor, Arabinose BAD Operon Promotor, Lambda-abgeleiteter P_L Promotor und N Gen Ribosomen Bindungsstelle; zudem der hybride tac Promotor aus den Sequenzen der Tryptophan und Lac UV5 Promotoren. Darunter fallen Hefepromotoren wie diejenigen für die 3-Phosphoglyzeratkinase, die Glyceraldehyd-3-Phosphat-Dehydrogenase, die Galaktokinase (GAL1), die Galaktoseepimerase, die saure Phosphatase (PHO5) und Alkohol-
40 dehydrogenase (ADH). Für höhere eukaryontische Zellen fallen darunter die Promotoren für den Simian Virus 40 (SV40), Rous Sarkoma Virus (RSV), Adenovirus (ADV), bovinem Papilloma Virus (BVPV) und CMV (Zytomegalievirus). Ebenfalls gehören dazu die entsprechenden Terminator- und Poly-A kodierenden Sequenzen, sowie Enhancer-Sequenzen, welche die Expression
45

erhoben können. Peptidsequenzen, welche die Sekretion der rekombinanten 3GnT erleichtern wie Signalsequenzen, Präprosequenzen usw. sind eingeschlossen und allgemein bekannt.

Die Erfindung betrifft auch Nukleinsäuren, welche für die Wildtyp-Form oder eine abgeleitete Form der 3GnT kodieren, und welche in die Zellen transfektiert werden und mit der endogenen

5 DNA der Zelle homolog oder nicht homolog rekombinieren

Die Erfindung betrifft auch Nukleinsäuren, welche als Sonden zur Entdeckung der 3GnT in anderen Spezies als den hier erwähnten und durch die Seq ID Nr 1a bzw 1b spezifizierten murinen und humanen Formen, gebraucht werden.

10 *Polypeptide und Antikörper*

Die Erfindung betrifft isolierte Peptide und Polypeptide, für welche die erfindungsgemässen Nukleinsäuresequenzen kodieren. Bevorzugt sind Peptide mit mindestens fünf Aminosäuren.

Nukleinsäuren, welche für Proteine kodieren, können zur Expression rekombinanter Polypeptide in intakten Zellen oder in vitro Translationssystemen verwendet werden. Die zur Expression ver-

15 wendeten Nukleinsäuren können neben den herkömmlichen molekularbiologischen Methoden auch chemisch synthetisiert werden. Die erfindungsgemässen Polypeptide mit den Seq-ID Nr. 2a bzw. 2b inklusive der Funktions-konservierten Varianten können aus natürlichen Organismen

bzw. Zellen, oder aus Organismen bzw. Zellen, welche die rekombinante 3GnT heterolog exprimieren, isoliert werden. Die Organismen sind beispielsweise transgene Mäuse, welche das 3GnT

20 Gen überexprimieren oder Nutztiere, welche das 3GnT Transgen in der Milchdrüse exprimieren. Die Zellen sind beispielsweise Bakterien, Pilze, Insekten, Pflanzen und tierische Zellen. Die heterolog exprimierten Polypeptide können auch als Fusionsproteine exprimiert werden.

Die Methoden zur Isolierung von Polypeptiden sind allgemein bekannt. Nicht begrenzend aufgezählt fallen darunter präparative Gel-Elektrophorese, präparative isoelektrische Fokussierung,

25 HPLC, FPLC, Gel Filtration, Ionenaustauschchromatographie, Aussalzen und Affinitätschromatographie. Zur erleichterten Isolierung der rekombinanten 3GnT kann das Enzym mit einer Bindungsstelle gekoppelt werden, wie eine Polyhistidin-Sequenz von 4-8 Histidinresten (sog. "His-

Tag", wobei die bevorzugte Variante 6 His-Reste enthält), oder Epitope für die Bindung von monoklonalen Antikörpern wie das FLAG-Epitop oder Myc-Epitop und andere. Bevorzugt ist

45 eine Immunaффinitätschromatographie mit Antikörpern gegen das rekombinante Enzym selber.

Das Enzym kann als Hybridenzym exprimiert werden, wobei es direkt an ein anderes Enzym gekoppelt wird zur Erleichterung der Katalyse. Bevorzugte Varianten sind mit einer

50 β 1,4Galactosyltransferase gekoppelte 3GnT zur Herstellung des Keratan- und Polylactosamingly-

5 cangerustes. Diese Hybridenzyme können entweder als Fusionsproteine oder als getrennte Proteine mit oder ohne "internal ribosomal entry sites" getrennte bicistronischen Vektoren exprimiert werden. Bevorzugte bicistronische Vektoren sind solche, bei denen die Reihenfolge der Expressi-
10 on zu einer Angleichung der katalytischen Aktivität der exprimierten Produkte führen. Solche Hybridenzyme können auch in multicistronischen Vektoren, welche noch für andere Proteine wie Resistenzmarker und/oder andere Glykosyltransferasen kodieren, enthalten sein.

Die erfindungsgemässen Polypeptide können im Sinne posttranslatorischer Modifikationen verändert werden, wie durch Phosphorylierung, Sulfatierung, Acylierung und Glycosylierung. Auch künstliche Veränderungen wie der Einbau radioaktiver Isotope beispielsweise durch metabolische
15 Markierung oder die Kopplung mit Fluoreszenzfarbstoffen sind eingeschlossen. Ebenso fallen darunter Fusionsproteine mit fluoreszierenden Proteinen wie dem "Green fluorescent protein".

Die Erfindung betrifft Antikörper, welche immunogene Komponenten der 3GnT erkennen. Diese Antikörper können zur spezifischen Bindung des Enzyms mit oder ohne Neutralisation seiner Aktivität eingesetzt werden. Unter den Antikörpern gegen die erfindungsgemässen Polypeptide fallen
20 sowohl mono- wie polyklonale Antikörper. Solche Antikörper werden durch in vivo Immunisierung von Tieren wie Mäuse, Ratten, Kaninchen, Ziegen, Esel und andere gewonnen oder durch in vitro Immunisierung von immunkompetenten Zellen. Die dazu verwendeten Antigene als Teile der 3GnT werden entweder aus natürlichen oder heterologen Quellen gewonnen. Die Antikörper selber können auch rekombinant hergestellt oder biochemisch aus schweren und leichten Ketten gebildet werden. Die Antikörper betreffen Hybridantikörper, bei welchen die Ketten mit zwei verschiedenen Spezifitäten gegen die 3GnT verbunden werden, univalente Antikörper und FAB sowie (FAB)₂ Fragmente. Die verschiedenen Antikörperformen sind in der Literatur beschrieben.
30 Ihre Reinigung erfolgt bevorzugt über Affinitätschromatographie; die erwähnten Methoden zur Polypeptidreinigung finden auch für die Reinigung von spezifischen Antikörpern Anwendung.
25 Die Antikörper können zur Quantifizierung der erfindungsgemässen Polypeptide angewendet werden. Die dazu verwendeten Methoden sind Enzyme-linked immunosorbent assays (ELISA) oder Radioimmunoassays (RIA) welche mit verschiedenen Enzymen/Substraten bzw. Isotopen durchgeführt werden. Diese Methoden sind in der Literatur beschrieben.

45 30 Die Erfindung wird durch nachstehende Beispiele illustriert:

5

Beispiele:

(Die Abbildungen befinden sich am Ende des Patentantrags)

10

*Beispiel 1:*Klonierung der 3GnT, einer mit der β 1,3GalT Genfamilie homologen cDNA

15

- 5 Mit Hilfe des tblastx Algorithmus wurden EST-Sequenzen, welche mit denjenigen der Mäuse β 1,3Galaktosyltransferase-Familie (Hennet T. et al., J. Biol. Chem. 1998, 273, 58-65) Ähnlichkeit aufweisen, in der EST-Sektion der GenBank identifiziert. Die menschliche und murine 3GnT cDNA wurde mit Hilfe eines 367 bp¹¹ Fragments, welches aus der EST Identifikations Nr AA150140 abgeleitet worden ist, aus einer menschlichen Fötalgehirnbank (CLONTECH) bzw.
- 10 einer Gehirnbank neugeborener Mäuse (Stratagene) isoliert. Die dazu benötigte Sonde wurde mit Hilfe der PCR¹² aus 50 ng menschlicher T Zell cDNA als Matrize mit folgendem Primerpaar 5'-GCGACTACTACCTGCCCTACG-3' und TCCCTTCTCTGGCAAGCACT-3' mit 30 Zyklen bei 95 °C für 45 s, 58 °C für 30 s und 72 °C für 45 s hergestellt.

20

25

15 *Beispiel 2*

Sequenzierung der 3GnT cDNA: Die humane und murine cDNA, welche wie im Beispiel 1 beschrieben aus den entsprechenden Genbanken isoliert worden ist, wurde gemäss der Dideoxy-Kertenterminationsmethode (Sanger F., Science 1981; 214,1205-1210) unter Benützung des Enzyms Sequenase 2.0 (USB) sequenziert. Die dazu benötigten Primer sind in der Tabelle 1 aufgeführt.

30

35

40

45

50

¹¹ bp steht für Basenpaare, kbp steht für kilo-Basenpaar

¹² PCR bezeichnet Polymerase Chain Reaction

55

Tabelle 1: Primer für

Humane 3GnT	Murine 3GnT
5'-CAGGAAACAGCTATGAC-3' (M13 rev)	5'-CAGGAAACAGCTATGAC-3' (M13 rev)
5'-GTAATACGACTCACTATAGGGC-3' (T7)	5'-GTAATACGACTCACTATAGGGC-3' (T7)
5'-CAATGGGTTTCAGTGACTTG-3'	5'-AACCTCACGGCCAAGGTCCTG-3'
5'-AGCGACCACAAGCGTGGCTG-3'	5'-GGATGCTTGGCCTGTGTCAG-3'
5'-AGGTAGTAGTCGCAGAGTTG-3'	5'-CAAGCATCCTGAACAATCTG-3'
5'-GTGAGGTTTTCTAGGCGTC-3'	5'-CATCCAGCCAGGTCAGCATG-3'

Die Sequenz der humanen 3GnT ist als Seq-ID Nr. 1a, diejenige der murinen als Seq-ID Nr. 1b vollständig beschrieben.

Beispiel 3

Proteinsequenzvergleich: Schema 4 vergleicht die Protein Sequenzen der Maus 3GnT mit den bekannten Maus β 3Gal-Ts, als sog. "ClustalW alignment. Konservierte Aminosäure werden mit schwarzem Hintergrund dargestellt. Die weisse Pfeile zeigen die Cysteine, welche unter den β 3Gal-Ts konserviert sind und der schwarzer Pfeil zeigt die einzige Cystein-Stelle, die in den fünf Proteinen konserviert ist.

Schema 4: Sequenzvergleiche der vier murinen homologen β 1,3Galactosyltransferasen mit der erfindungsgemässen murinen 3GnT (entsprechend Seq-ID Nr. 2b)

	20	40	60	80	
β 3GalT-I :	-----	MASKVSLVYLSVVCNASALWYL	-----		: 23
β 3GalT-II :	MLQWRRRHCCFAKMTWSPKRSLLRTP	LTGVLSLVFLFAMFLFFNHDM	LPGPGFKNPVYTFRGFRSTKSETHSSLR		: 80
β 3GalT-III :	-----	MAPAVLTALPNRMSLRSLKWSLLLSLSFL	-----		: 31
β 3GalT-IV :	-----	MPLSLFRVLLAVLLVITLTLFG	-----		: 24
β 3GnT :	-----		-----		: -
	100	120	140	↓	160
β 3GalT-I :	-----	SITRPTS	-----	SYTGSKP	-----
β 3GalT-II :	-----	TIWKEVAPQTLRPHIASNSNT	ELSPQGVTLQNTLSANGSIYNEKGTGHPNSYHFKYIINEPEKCKEKSQ	-----	: 83
β 3GalT-III :	-----	VIWYLSLP	-----	HYNVIERVNMYFYEY	-----
β 3GalT-IV :	-----	PSGLG	-----	EELLSLSLASLLPAPASGPPPLALPRLLISNSH	-----
β 3GnT :	-----	MKVFR	-----	AWRHRVALGLGLAFCGT	-----
	180	200	220	240	
β 3GalT-I :	ISTTHKEFCA	ROATREIGDENNFKE	IKIATLFLGKNADP	-----	VLNQWEDSSQIFHO
β 3GalT-II :	IAAEPGQIEA	RAIRQITGNETLAPG	IQIIRVLLGISIKLNG	YLQHAQESRQYHO	LIQQEYLDITVYNIITKTLN
β 3GalT-III :	VTSRPSDVKA	QATRMVTEGKSWWC	YEVLTFFLLGQQAEREDKT	LALSLED	QVLYGO
β 3GalT-IV :	YCTAPEHLHQ	WATRASHGAIREARQ	FRVOTLFLGKPRRQ	-----	CLADSS
β 3GnT :	MASAPRAVER	INTAVRSTLAPERRG	CPEDVNARFAVCTGGLGS	-----	EERRALEL
	260	280	300	320	
β 3GalT-I :	GMRVATFCSKAKYV	HTQSTIRVNMDNLYKLLKPS	TKPRRR	-----	YFTGWIAG
β 3GalT-II :	GNNVATYCPT	PNVHKIOSDFVNTLEYL	KLLKPDLP	PRHN	-----
β 3GalT-III :	AFRVMFECPN	ARYITXITDQVHINTGML	KYLLNLHSEK	-----	FFTPPLIDNYSY
β 3GalT-IV :	GLNWNKYCP	MARYITLITDQVYVNP	ELVSELIQ	RGGPSQWQKGEAQEET	TAIHEEHRGQAVPL
β 3GnT :	MLTHIDERV	DFEVLKADDS	PEARLDAHVD	RAREPARRRR	-----
	340	360	380	400	
β 3GalT-I :	ROVRSKWY	PPDLTPDSN	-----	YPPFCS	STGYIFADVAEL
β 3GalT-II :	FNKDSKWY	PPDLTPSER	-----	YPPVCS	STGYVFGDLAEK
β 3GalT-III :	RGFFHNKHS	YQEPFKV	-----	FPYCS	GLGYMGGDLVPR
β 3GalT-IV :	RTPESRHHV	SEELPENWGP	PPPYAS	STGYVLSISAVQL	LKVASRAPP
β 3GnT :	RVKPG	GRAREAAQLCD	YLLPYALGGVLSADLVHY	RLRSREYLRAWHSE	DVSLGTWLPVDVQREHDP
	420	440	↓	460	480
β 3GalT-I :	KMAYSLGRYR	ATVYQISPEE	HRINDDSSKK	-----	H
β 3GalT-II :	RVSYSCKYSH	LITSQFQPS	ELIKYNNHQQNK	-----	H
β 3GalT-III :	RIHLDVQL	RRVIAAGFSSKE	LITFAQVLRN	-----	TTCHY
β 3GalT-IV :	PLDR	COYGKFLITS	IKVDPWQ	QEAQKLYSGMNGERTA	-----
β 3GnT :	YKSR	GNNYQLVLT	IKQSPED	LEKQMLHEGRLCKHEVQLRLSYVVDW	SAPPSCCQKKEGVP

5 Beispiel 4

Die Expression der 3GnT im Organismus: Die exprimierten mRNAs wurden mit Hilfe der Northern blot Analyse unter Anwendung kommerziell erhältlicher Analysefilter (englisch "blot"), auf welchen die elektrophoretisch aufgetrennten mRNA's verschiedener Organe immobilisiert sind (CLONTECH), durchgeführt. Um DNA-Sonden mit erhöhtem GC-Gehalt zu vermeiden, wurden

10 3'-Fragmente der murinen und humanen 3GnT präpariert. Ein *SacI-PstI* 668 bp-Fragment der

5 murinen β 3GnT cDNA und ein 367 bp Fragment der Region der humanen β 3GnT cDNA zwischen Nukleotid 616 and 977 wurden mit [α - 32 P]-CTP phosphoryliert (Hartmann Analytics, Braunschweig, Germany) mit "random priming" (Sambrook et al., 1989, Molecular Cloning: A
10 laboratory Manual, 2. Auflage, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) und mit den Poly(A)⁺ RNA Filter (blots) hybridisiert. Die Filter (blots) wurden in 0,1 x SSC, 0,1% SDS bis zu 55 °C gewaschen and für vier Tage in Verstärkerkassetten bei -70 °C. exponiert.

15
20
25
30
35
40
45
50
Abb. 1: Expression der erfindungsgemässen 3GnT in verschiedenen Geweben der Maus und des Menschen (Northern Blot Analyse)

Die Abb. 1 zeigt das Resultat: Jede Spur zeigt ca 2 μ g Poly(A)⁺ RNA. Links sind die RNA Grössenmarker in kb angegeben. In allen untersuchten menschlichen Organen wurden drei Transkripte (1,6; 2,4; und 3,3 kb als stärkstes Signal) gefunden. Bei der Maus wurde ein 2,2 kb Transkript in den meisten Geweben gefunden, dazu eine schwächer exprimierte Form von 3,7 kb in der Lunge und Niere. Schwache Signale fand man in der Leber und im Skelettmuskel.

Beispiel 5

Expression der 3GnT: Ein *EcoRI-XhoI* 1,9 kbp Fragment wurde aus einer menschlichen 3GnT cDNA, welche für ein Protein entsprechend der Seq ID Nr 3 kodiert, abgeleitet und im FastBac-HTc Vektor (Life Technologies), welcher mit *EcoRI-XhoI* vorher linearisiert worden ist, subkloniert. Das Maus *EcoRI-XhoI* Gen wurde als *StuI-StuI* 1 kbp Fragment, welches der kodierenden Region ohne cytoplasmatische und Transmembrandomäne entspricht, im FastBac-HTc Vektor subkloniert. Rekombinante Baculoviren wurden mit der Transposon-vermittelten Rekombination entsprechend des vom Hersteller empfohlenen Protokolls hergestellt.

Die Insekten-Zellen wurden mit einer MOI¹³ von 10 infiziert und bei 27 °C wachsen gelassen bis zur Messung der 3GnT Aktivität.

¹³ MOI bezeichnet "multiplicity of infection"

Beispiel 6**Messung der Aktivität der rekombinanten 3GnT:**

5 x 10⁶ Insekten-Zellen, welche mit dem Wildtyp oder mit dem rekombinanten Baculovirus infiziert worden waren, wurden 72 h nach Infektion mit 2% Triton X-100 in 1 ml Phosphat-Puffer während 15 min auf Eis lysiert. Das Lysat wurde in Gegenwart von UDP-¹⁴C-GlcNAc als Donor- und LacNAc-pNP¹⁴ (Toronto Research) als Akzeptorsubstrat während 60 min bei 37 °C unter Zugabe folgender Ingredienzien inkubiert: 10 µl des Zell-Lysates in 50 µl eines 50 mM Kakodylat Puffers, pH 7.0, 20 mM MnCl₂, 5% Me₂SO, 0.75 mM ATP, 0.5 mM UDP-GlcNAc mit Zugabe von 5 x 10⁴ cpm von UDP-[¹⁴C]GlcNAc (Amersham) und verschiedenen Akzeptorsubstraten. Die Reaktion wurde gestoppt durch Verdünnung mit 0,4 ml eiskaltem H₂O. Das Enzymgemisch wurde auf eine SeP-Pak Säule gegeben, gewaschen und das haftende Produkt mit Methanol eluiert und mit Flüssigscintillationszählung gemessen. Folgende Resultate wurden erhoben (siehe Tabelle 2):

Tabelle 2: Substratspezifität der erfindungsgemässen 3GnT

Akzeptoren		Sf9 ^a Wildtyp	3GnT ^b
		pmol/min mg prot	
Galβ1-4GlcNAc-pNP ^c	5 mM	6	1,478
Galβ1-4Glc-bzl ^d	5 mM	4	1,034
Galβ1-4GlcNAc-octyl ^e	2 mM	5	284
Galβ1-3GlcNAc-octyl	2 mM	5	8
Galβ1-3GalNAc-octyl	2 mM	5	4
Galα-pNP	5 mM	4	89
Galβ-pNP	5 mM	4	26
GalNAcα-bzl	5 mM	4	5
GalNAcβ-bzl	5 mM	5	6
GlcNAcα-bzl	5 mM	4	4
GlcNAcβ-bzl	5 mM	11	7
GlcNAcβ1-3GalNAc-bzl	5 mM	8	5

^a: Sf9 steht für *Spodoptera frugiperda* 9-Zell-Lysat mit Wildtyp Baculovirus infiziert, als neg. Kontrolle

^b: Sf9 Zell-Lysat (mit 3GnT-rekombinantem Baculovirus infiziert)

^c: pNP ist *para*-nitrophenyl

^d: bzl ist benzyl

^e: octyl, O-(CH₂)₈-CO₂Me

¹⁴ LacNAc-pNP bezeichnet N-azetyl-laktosamin-p-nitrophenyl

*Beispiel 7*Polylactosaminsynthase-Aktivität der 3GnT

Das Enzym wurde im FastBac-HTc vector (Life Technologies) subkloniert, in 5×10^6 Sf9 Zellen transfiziert und 72 h nach Infektion in 1 ml eines 2% Triton X-100-haltigen Phosphatpuffers (pH 7,4), auf Eis während 15 min lysiert. Die zytoplasmatische und solubilisierbare Membranfraktion wurde mit Zentrifugation bei $300 \times g$ während 5 min gewonnen. Die $\beta 3$ GnT Aktivität wurde mit 10 μ l des Zelllysates in 50 μ l eines 50 mM Kakodylat Puffers, pH 7,0, 20 mM $MnCl_2$, 5% Me_2SO , 0,75 mM ATP, 0,5 mM UDP-GlcNAc mit Zugabe von 5×10^4 cpm von UDP- $[^{14}C]$ GlcNAc (Amersham) und verschiedenen Akzeptorsubstraten gemessen. Die Reaktion wurde gestoppt durch Zugabe von 0,5 ml eiskalten Wassers. Die Reaktionsprodukte wurden von den Edukten durch hydrophobe Chromatographie an Sep-Pak C_{18} Säulen (Waters) getrennt im Falle des Gebrauch von Akzeptoren mit hydrophobem Aglycon (Hennet T. et al., J. Biol. Chem. 1998, 273, 58-65). Im Falle der Anwendung von Laktose, *N*-Azetyllaktosamin, Gal- $\beta 1,3$ -GlcNAc-*N*-Azetyllaktosamin (lac-*N*-tetraose) und Gal- $\beta 1,4$ -GlcNAc-*N*-Azetyllaktosamin (lac-*N*-neo-tetraose), wurden die Reaktionsprodukte an AG1-X8 Säulen (Bio-Rad) gereinigt, indem 2 ausgeschlossene Volumina gesammelt wurden (2 ml) (Malissard, M., et al., Eur. J. Biochem. 1996, 239, 340-348). Die Hydrolyse von UDP-GlcNAc in Zell-Lysaten wurde in Enzymansätzen ohne Akzeptorsubstrate gemessen. Die $[^{14}C]$ -markierten Produkte wurden mit Flüssigszintillationszählung gemessen (Rackbeta, Pharmacia). Die Resultate sind auf Abb. 6 aufgeführt.

Abb. 2: Polylactosaminsynthase-Aktivität der erfindungsgemässen 3GnT

Die Abb. 2 zeigt die Eigenschaft der erfindungsgemässen 3GnT, den Tetrasaccharid- Akzeptor lac-*N*-neo-tetraose effektiv umzusetzen. Dieser Befund beweist, dass das 3GnT Enzym Polylaktosaminketten sowohl initiieren wie auch verlängern kann.

*Beispiel 8*Identifikation des erfindungsgemässen 3GnT-Produktes mit Massenspektrometrie und Methylierungsanalyse

Um den Bindungstyp der von der erfindungsgemässen 3GnT katalysierten glykosidischen Bindung zu bestimmen, wurde eine Inkubation mit Gal($\beta 1-4$)Glc($\beta 1$ -OBzl) al Akzeptor im grossen Massstab durchgeführt. Die Inkubationsbedingungen entsprachen denjenigen unter Beispiel 6. Das Trisaccharidprodukt wurde mit HPLC isoliert und mit verschiedenen Methoden analysiert. Seine

5 molekulare Masse wurde nach Permethylierung mit dem positiv-Ionen Modus MALDI TOF Massen Spektrometrie bestimmt. Die erhobene Bruttoformel war $\text{HexNAcHex}_2\text{Bn}$ (m/z 798, $[(M-\text{Na})^+]$). Die Methylierungsanalyse zeigte die Gegenwart von 4'-substituierten Glc, 3'-substituierten Gal, und terminales GlcNAc; damit konnte die (1-3)-Bindung zwischen GlcNAc
10 und Gal bewiesen werden.

Beispiel 9

Identifikation des erfindungsgemässen 3GnT-Produktes mit Magnetresonanzspektrometrie¹⁵

Um die Richtigkeit der unter Beispiel 7 beschriebenen Analyse zu bestätigen und um die anomere
10 Konfiguration des von der erfindungsgemässen 3GnT übertragenen Zuckers zu bestimmen, wurden 1D und 2D ^1H NMR Experimente durchgeführt. Die am meisten distal verlagerten Signale im 1D ^1H NMR Spektrum (Fig. 3a) des Trisaccharids gehörten zur Benzyl Gruppe (δ 7.431, 4.930 (J_{gem} 11.6 Hz), 4.755 (J_{gem} 11.6 Hz). Mit Hilfe von 2D TOCSY (20 und 100 ms) und ROESY Experimenten konnten die drei anomeren Doppelsignale bei δ 4.673 ($^3J_{1,2}$ 8.6 Hz), 4.546 ($^3J_{1,2}$ 8.0
25 Hz), und 4.423 ($^3J_{1,2}$ 8.0 Hz) den Strukturen β -GlcNAc, β -Glc, bzw β -Gal, zugeordnet werden. Das ROESY Spektrum (200 ms, Fig. 3b) erlaubte die Bestimmung der zwei glykosidischen Bindungen. Die anomere Spur des Gal Restes zeigte neben zwei innermolekularen "Crosspeaks" (Gal H-1,H-3, δ 3.71; Gal H-1,H-2 (TOCSY Transfer), δ 3.573), einen intermolekularen *crosspeak* zwischen Gal H-1 und Glc H-4 (δ 3.639). Die anomere Spur des Glc Restes zeigte drei intramolekulare *crosspeaks* bei δ 3.604 (Glc H-1,H-3), 3.573 (Glc H-1,H-5), und 3.338 (Glc H-1,H-2
30 (TOCSY Transfer)). Schliesslich zeigte die anomere Spur des GlcNAc Restes neben drei intramolekularen *crosspeaks* (GlcNAc H-1,H-3, δ 3.557, GlcNAc H-1,H-5, δ 3.446, GlcNAc H-1,H-2 (TOCSY Transfer), δ 3.74), einen intermolekularen *crosspeak* zwischen GlcNAc H-1 und Gal H-3 (δ 3.71). Zusammenfassend ergeben die Resultate den Nachweis des Produktes GlcNAc(β 1-3)Gal(β 1-4)Glc(β 1-OBn).
40

Abb. 3: 600-MHz 1D (a) und 2D ROESY (200ms) (b) ^1H NMR Spektrum des isolierten Trisaccharids GlcNAc(β 1-3)Gal(β 1-4)Glc(β 1-OBn) (C-B-A).

¹⁵ Die Magnetresonanzspektrometrie wird in diesem Schriftstück als NMR (Nuclear magnetic resonance) bezeichnet

Claims

5

10

15

20

25

30

35

40

45

50

55

5

Patentansprüche

10

1. Eine isolierte Nukleinsäure, welche für die UDP-*N*-Azetylglukosaminyl- β 1,4 Galaktosid:
 β 1,3*N*-Azetylglukosaminyltransferase (Polylaktosaminyl Typ) kodiert, gemäss Seq-ID Nr. 1a bzw

5

1b.

15

2. Eine von der Nukleinsäure nach Anspruch 1 abgeleitete Nukleinsäure.

3. Ein DNA-Konstrukt, das eine DNA-Sequenz gemäss Anspruch 1 enthält.

4. Ein DNA-Konstrukt, das eine DNA-Sequenz gemäss Anspruch 2 enthält.

5. Eine Zelle, welche ein DNA-Konstrukt nach Anspruch 3 enthält.

10

6. Eine Zelle, welche ein DNA-Konstrukt nach Anspruch 4 enthält.

20

7. Ein Verfahren zur Herstellung von 3GnT-Polypeptiden dadurch gekennzeichnet, dass

a) ein isoliertes DNA-Molekül, welches für die 3GnT kodiert oder ein davon abgeleitetes Kon-
strukt in eine Wirtszelle eingeführt wird,

b) die Wirtszelle unter die für die Expression von 3GnT günstigen Bedingungen kultiviert wird

25

15

c) die 3GnT, welche von den Wirtszellen produziert wird, isoliert wird.

8. Ein Verfahren für die Herstellung von Di-oder Oligosacchariden dadurch gekennzeichnet, a)
dass die genannten Di-oder Oligosaccharide die Struktur $\text{GlcNAc}\beta 1\Rightarrow 3\text{Gal}$ enthalten

30

b) dass enzymatisch aktive, nach Anspruch 7 hergestellte 3GnT UDP-GlcNAc als Donorsubstrat
und β Gal-terminierte Akzeptorsubstrate erkennt und den GlcNAc-Rest überträgt.

20

9. Ein Verfahren nach Anspruch 8 dadurch gekennzeichnet, dass Glykanketten beliebiger Länge,
welche das Disaccharid $\text{GlcNAc}\beta 1\Rightarrow 3\text{Gal}$ als repetitives Element enthalten, in vitro hergestellt
werden, indem dem Reaktionsansatz zusätzlich eine β 1,4Galaktosyltransferase mit UDP-
Galaktose als Donorsubstrat zugegeben wird.

35

25

10. Ein Verfahren nach Anspruch 9, dadurch gekennzeichnet, dass Glykanketten beliebiger Länge,
welche das Disaccharid $\text{GlcNAc}\beta 1\Rightarrow 3\text{Gal}$ als repetitives Element enthalten, in vivo hergestellt
werden, indem die Zelle zusätzlich mit einer β 1,4Galaktosyltransferase transfektiert wird.

40

45

50

ERSATZBLATT (REGEL 26)

55

Abb. 1

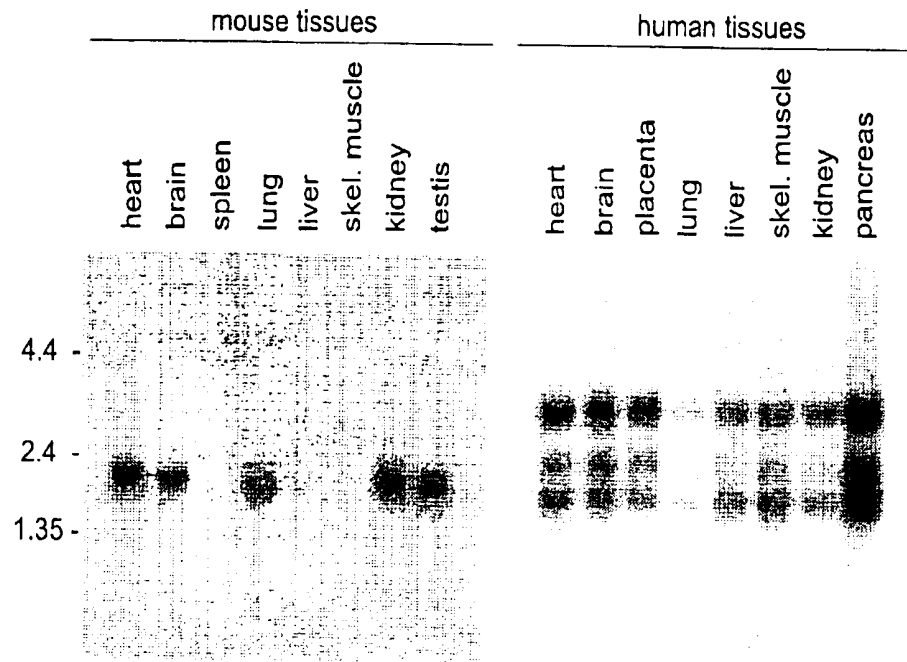


Abb. 2

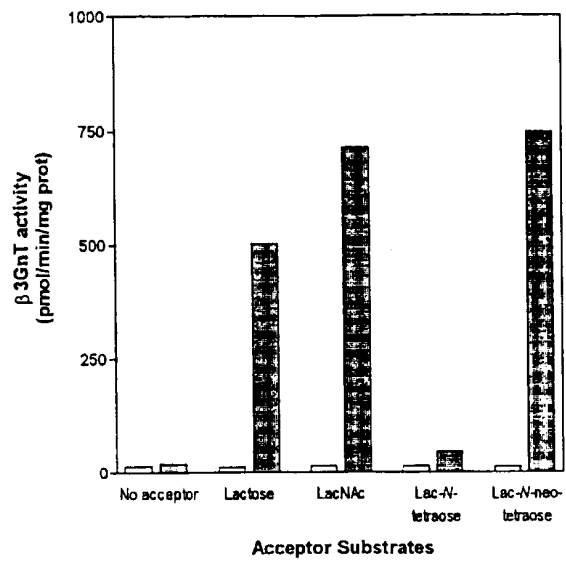
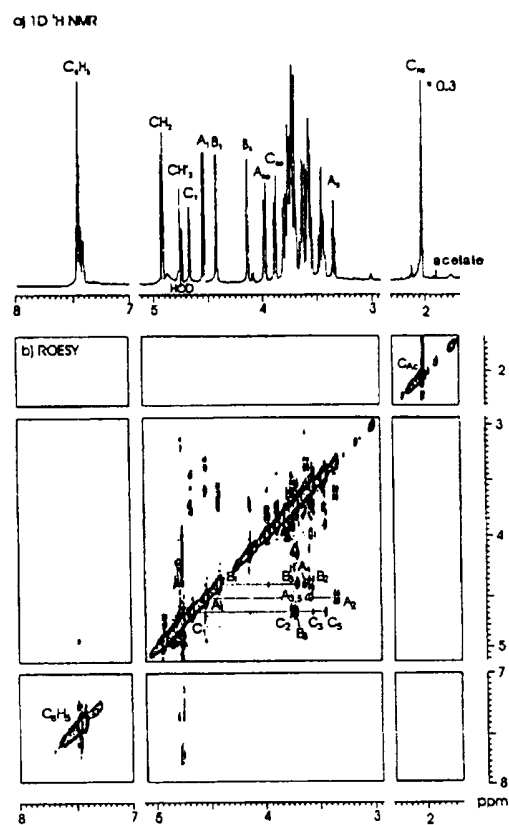


Abb. 3



Die folgenden Schematas und detaillierten Ausführungen ergänzen die Beschreibung der Erfindung

Schema 1a: Verbindung mit der Sequenzidentifikationsnummer 1a

- 5 Dieses Schema zeigt die Nukleinsäuresequenz der cDNA, welche für die humane UDP-N-Azetylglukosaminyl- β 1,4 Galaktosid: β 1,3N-Azetylglukosaminyltransferase (Polylaktosaminyl Typ) kodiert
- (1) Seq-ID Nr. 1a.
- (i) Sequenzkennzeichen:
- 10 (A) LÄNGE: 990 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: EINZELSTRANG
- (D) TOPOLOGIE: linear
- 15 (ii) ART DES MOLEKÜLS: cDNA zu mRNA
- (iii) URSPRÜNGLICHE HERKUNFT:
- (A) ORGANISMUS: Homo sapiens
- (D) ENTWICKLUNGSSTADIUM: Fötalstadium
- 20 (F) GEWEBETYP: Gehirn
- (vii) UNMITTELBARE HERKUNFT:
- (A) BIBLIOTHEK: λ -Phagen Genbank aus fötalem menschlichen Gehirn (CLONTECH, Palo
- 25 Alto, CA, USA)
- (B) KLONE: GT99
- (ix) MERKMAL
- 30 (A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für cytoplasmatische Domäne
- (B) LÄGE: 1-33

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für Transmembran-Domäne

(B) LAGE: 34-90

5

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für Stamm und Katalytische Domäne

(B) LAGE: 91-990

10

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für das Gesamtprotein

(B) LAGE: 1-990

15

	1	ATGAATCTGC	TGCGGCGGGC	GTGGCGGCGG	CGGGCGGCGC	TAGGCCTGGG	CACGCTGGCG
		TACTTAGACG	ACGCGCGCCG	CACGCGCGCC	GCCCGCGCGG	ATCCGGACCC	GTGCGACCGC
	61	CTGTGCGGGG	CGGCGCTGCT	CTACCTGGCG	CGCTGCGCGG	CCGAGCCCGG	GGACCCCAAG
20		GACACGCCCC	GCCGCGACGA	GATGGACCGC	GCGACGCGCC	GGCTCGGGCC	CCTGGGGTCC
	121	GCGATGTCGG	GCCGCGAGCC	GCCTCCCCCC	GCGCCCGCGC	GCGCCGCGCG	CTTCGTGGCA
		CGCTACAGCC	CGGCGTCGGG	CGGAGGGGGG	GCGGGGCGCG	GCGGGGCGCG	GAAGGACCGT
25	181	GTGCTGGTGG	CCAGCGCGCC	CCGCGCGCGC	GAGCGCGCGA	GCGTGATCCG	CAGCAGCTGG
		CACGACCAAC	GGTCGCGCGG	GCGCGCGCGG	CTCGCGCGGT	CGCACTAGGC	GTCTGTGCAC
	241	CTTGCGCGGG	GCGGGGCGCC	GGGCGACGTG	TGGGCGCGCT	TTGCCGTGGG	CACGGCGCGC
		GAACGCGCGG	GCGCCCGGGG	CCCGCTGCAC	ACCGCGCGGA	AACGGCACCC	GTGCCGGCGG
30	301	CTGGGCGCGG	AGGAGCGGGC	CGCCCTGGAG	CGGGAGCAGG	GCGGGCACGG	GGACCTGCTG
		GACCCGCGGG	TCCTCGCCGC	GCGGGACCTC	GCCCTCGTCC	GCGCCGTGCC	CCTGGACGAC
	361	CTGTGCGCGG	CGCTGCGCGA	CGCCTACGAA	AACCTCACGG	CCAAGGTGCT	GGCCATGCTG
35		GACGACCGGG	GCGACGCGCT	GCGGATGCTT	TTGGAGTGCC	GGTTCACGGA	CCGGTACGAC
	421	GCCTGGCTGG	ACGAGCACGT	GGCCTTCGAG	TTGCTGCTCA	AGGCGGACGA	CGACTCCTTC
		CGGACCGACC	TGCTCGTGCA	CCGGAAGCTC	AAGCACGAGT	TCCGCCTGCT	GCTGAGGAG
40	481	GCGCGGCTGG	ACGCGCTGCT	GGCCGAGCTG	GCGCGCCCGG	AGCCCGCGCG	CCGCCGCGCG
		CGCGCCGACC	TGCGCGACGA	CCGGCTCGAC	GCGCGGGCGC	TCGGGCGCGC	GCGCGCGCGG
	541	CTCTACTGGG	GCTTCTTCTC	GGGCGCGCGC	CGAGTCAAGC	CGGGGGGGCG	CTGGCGCGAG
45		GAGATGACCC	CGAAGAAGAG	CCCGGCGCGG	GCTCAGTTCC	GCCCCCGCGC	GACCGCGCTC
	601	GCGCGCTGGC	AACTCTGCGA	CTACTACCTG	CCCTACGCGC	TGGGCGGGCG	CTACGTGCTC
		CGGCGGACCG	TTGAGACGCT	GATGATGAC	GGGATGCGCG	ACCCGCGCGC	GATGCACGAG
50	661	TGGGCGGACC	TGGTGCACTA	CCTGCGCCTC	AGCCGCGACT	ACCTGCGCGC	CTGGCACAGC
		AGCCGGCTGG	ACCACGTGAT	GGACGCGGAG	TGGGCGCTGA	TGGACGCGCG	GACCGTGTCC

721 GAGGA-CTGT CTCTGGGCGC CTGGCTGGCG CCGGTGGACG TCCAGCGGGA GCACGACCCG
 CTCCTGACCA GAGACCCGCG GACCGACCGC GGCCACCTGC AGGTCGCCCT CGTGCTGGGC
 5 781 CGCTTCGACA CCGAATACCG GTCCCGCGGC TGCAGCAACC AGTACCTGGT GACGCACAAG
 GCGAASCTGT GGCTTATGGC CAGGGCGCGG ACSTCGTTGG TCATGGACCA CTGCGTGTTT
 841 CAGAGCTGGS AGGACATGCT GGAGAAGCAC GCGACGCTGG CGCGCGAGGG CCGCCTGTGC
 GTCTCCTACC TCCTGTACGA CCTCTTCGTG CGCTGCGACC GCGCGCTCCC GCGCGACACG
 10 901 AAGCGGAGG TGCAGCTGCG CCTGTCCTAC GTGTACGACT GGTCCGCGCC GCCCTCGCAG
 TTCGCGCTCC ACGTCGACGC GGACAGGATG CACATGCTGA CCAGGCGCGG CCGGAGCGTC
 961 TGCTGTCAGA GAAGGGAGGG CATCCCTGA
 ACGACCTCTT CTTCCCTCCC GTAGGGGACT
 15

Schema 1b: Verbindung mit der Sequenzidentifikationsnummer 1b

Dieses Schema zeigt die Nukleinsäuresequenz der cDNA, welche für die murine UDP-N-Azetylglukosaminyl- β 1,4 Galaktosid: β 1,3N-Azetylglukosaminyltransferase (Polylaktosaminyl

20 Typ) kodiert

(1) Seq-ID Nr. 1b.

(i) Sequenzkennzeichen:

(A) LÄNGE: 978 Basenpaare

(B) ART: Nukleinsäure

25 (C) STRANGFORM: EINZELSTRANG

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA zu mRNA

30 (iii) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Mus musculus

(D) ENTWICKLUNGSSTADIUM: Fötalstadium

(F) GEWEBETYP: Gehirn

35 (vii) UNMITTELBARE HERKUNFT:

(A) BIBLIOTHEK: λ -Phagen Genbank aus fötalem murinen Gehirn (Stratagene: La Jolla, CA, USA)

40 (B) KLONE Bezeichnung wählen

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für cytoplasmatische Domäne

5 (B) LAGE: 1-33

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für Transmembran-Domäne

10 (B) LAGE: 34-90

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für Stamm und Katalytische Domäne

15 (B) LAGE: 91-978

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Sequenzabschnitt codierend für das gesamte Protein

20 (B) LAGE: 1-978

1 ATGAAGGTAT TCCGGCGCGC TTGGCGGCAC CGGGTGGCGC TGGGCCTAGG CGGCCTGSCG
 TACTTCCATA AGGCCGCGCG AACCGCGTG GCCCACC GCG ACCCGGATCC GCCGACCGC
 25 61 TTCTGCGGCA CCACTCTGTT GTACCTGGCG CSETGCGCTT CCGAGGGCGA GACGCCCTCC
 AAGACGCCGT GGTGAGACAA CATGGACCGC GCGACGCGAA GGCTCCCGCT CTGCGGGAGG
 121 GCTTCGGGAG CCGCTCGGCC CCGCGCTAAG GCCTTCCTGG CGTGCTGGT GGCCAGTGCG
 CGAATGCCCT GGCAGGCCGG GCGCGGATTC CGGAAGGACC GCCACGACCA CCGGTCACGC
 30 181 CCCCSCGCGG TCGAGCGCGC CACCGCAGTG CGCAGCACGT GGCTGGCACC GSAGAGGCGT
 GGGGCGCGCC AGCTCGCGGC GTGGCGTCAC GCGTCGTGCA CCGACCGTGG CCTCTCCSCA
 241 GCGCGACCCG AGGACGTGTG GCGCGGCTTC GCCGTGGGCA CTGGCGGCTT AGGCTCGGAG
 35 CCGCCTGGGC TCCTGCACAC CCGCGCGAAG GCGCACCCGT GACCGCGGAA TCCGAGCCTC
 301 GAGCSCGCGC CTCTTGAGCT CGAGCAGGCG CAGCACGGGG ACCTGCTGCT GCTGCCCGCC
 CTCGCGCGGC GAGAAGTCGA GCTCGTCCGC GTCGTGCCCC TGGACGACGA CGACGGGCGG
 40 361 CTGCGCGACG CCTACGAGAA CCTCACGGCC AAGGTCCTGG CCATGCTGAC CTGGCTGGAT
 GACGCGCTGC GGATGCTCTT GGAGTGCCGG TTCCAGGACC GGTACGACTG GACCGACCTA
 421 GAGCSCGTGG ACTTCGAGTT CGTGCTCAAG GCGGACGACG ACTCCTTTGC GCGCCTGSCA
 CTCGCGCACC TGAAGCTCAA GCACGAGTTC CGCCTGTGCG TGAGGAAACG CCGGACCTG
 45


```

431 GGTATCCTGG TGGACCTACG CGCACGGGAG CCCGCACGCC GCCGGCGCCT CTACTGGGGC
CGATAGGACC ACCTGGATGC GCGTGCCCTC GGGCGTGCGG CGGCCGCGGA GATGACCCCG
5 541 TTCTTTTCCG GCGCGGGGCG CGTCAAGCCG GGAGGTGCGT GGCSAGAAAG AGCCTGGCAA
AAGAAAAGGC CCGCGCCCGC GCAGTTCGGC CCTCCAGCGA CCGCTCTTCG TCGGACCGTT
601 CTCTGGGACT ACTACCTGCC CTACGCGTTG GCGGTGGGCT ATGTCTTTTC TGCGGACCTG
GAGACGCTGA TGATGGACGG GATGCGCAAC CCGCCACCGA TACAGGAAAG ACGCCTGGAC
10 661 GTGCATTACC TGCGCCTCAG CCGCGAGTAC CTGCGCGCGT GGCACAGTGA AGACGTATCG
CACGTAATGG ACGCGGAGTC GCGGCTCATG GACGCGCGCA CCGTGTCAC TCTGCATAGC
721 CTGGGCACCT GGCTGGCACC AGTGGATGTG CAACGGGAGC ACGACCCACG CTTGCACACG
GACCCGTGGA CCGACCGTGG TCACCTACAC GTTGCCCTCG TGCTGGGTGC GAAGCTGTGC
15 781 GAGTACAAAT CTCGAGGCTG CAACAATCAG TATCTGGTGA CACACAAGCA AAGCCCAGAG
CTCATGTTTA GAGCTCCGAC GTTGTTAGTC ATAGACCACT GTGTGTTCGT TTCGGGTCTC
841 GACATGTTGG AGAAGCAACA GATGTTGCTG CATGAGGGCC GGTGTGTCAA GCATGAGGTG
20 CTGTACAACC TCTTCGTTGT CTACAACGAC GTACTCCCGG CCAACACGTT CGTACTCCAC
901 CAGTTGCGCC TTCTCTATGT CTATGACTGG TCAGCTCCAC CCTCCCAGTG CTGCCAGCGC
GTCAACGCGG AAAGGATACA GATACTGACC AGTCGAGGTG GGAGGGTCAC GACGGTCCCG
25 961 AAGGAGGGCG TTCCCTGA
TTCTCCCGC AAGGGACT

```

30 Schema 2a: Verbindung mit der Sequenzidentifikationsnummer 2a

Dieses Schema zeigt die Aminosäuresequenz der humanen 3GnT im Ein-Buchstabencode, welche durch Translation aus der Sequenz gemäss Seq ID-Nr. 1a übersetzt worden ist.

(1) Seq-ID Nr. 2a.

(i) Sequenzkennzeichen:

35 (A) LÄNGE: 329 Aminosäuren

(B) ART: Polypeptid

(D) TOPOLOGIE: linear

40 (ii) ART DES MOLEKÜLS: Polypeptid zu cDNA

(iii) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Homo sapiens

(D) ENTWICKLUNGSSTADIUM: Fötalstadium

45 (F) GEWEBETYP: Gehirn

(vii) UNMITTELBARE HERKUNFT:

(A) BIBLIOTHEK: λ -Phagen Genbank aus fötalem menschlichen Gehirn (CLONTECH, Adresse siehe oben), durch Translation bei heterologer Expression

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Cytoplasmatische Domäne

(B) LAGE: 1-11

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Transmembran-Domäne

(B) LAGE: 12-30

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Stamm und Katalytische Domäne

(B) LAGE: 31-329

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: N-Glycosylierungsstelle

(B) LAGE: Asn 131

(ix) MERKMAL

(A) NAME/SCHLÜSSEL: Gesamtprotein

(B) LAGE: 1-329

1 MNLRRRAWRR RAALGLGTLA LCGAALLYLA RCAAEPGDPR
41 AMSGRSPPPP APARAAFLA VLVASAPRAA ERRSVIRSTW
81 LARRGAPGDV WARFAVG TAG LGABERRALE REQARHGDL
121 LLPALRDAYE NLTA KVLAML AWLDEHVA FE FVLKADDDSF
161 ARLDALLAEL RAREPARRRR LYWGFFSGRG RVKPGGRWRE
201 AAWQLCDYYL PYALGGGYVL SADLVHYLRL SRDYLRWHS

241 EDVSLGAWLA PVDVQREHDP RFDTEYRSRG CSNQYLVTHK
281 QSLDMLEKH ATLAREGRLC KREVQLRLSY VYDWSAPPSQ
321 CCQRREGIP

5

Schema 2b: Verbindung mit der Sequenzidentifikationsnummer 2b

Dieses Schema zeigt die Aminosäuresequenz der murinen 3GnT im Ein-Buchstabencode, welche durch Translation aus der Sequenz gemäss Seq ID-Nr. 1b übersetzt worden ist.

10 (1) Seq-ID Nr. 2b.

(i) Sequenzkennzeichen:

(A) LÄNGE: 325 Aminosäuren

(B) ART: Polypeptid

15 (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Polypeptid zu cDNA

(iii) URSPRÜNGLICHE HERKUNFT:

20 (A) ORGANISMUS: Mus musculus

(D) ENTWICKLUNGSSTADIUM: Fötalstadium

(F) GEWEBETYP: Gehirn

(vii) UNMITTELBARE HERKUNFT:

25

(A) BIBLIOTHEK: λ -Phagen Genbank aus fötalem murinem Gehirn (Stratagene, Adresse siehe oben), durch Translation bei heterologer Expression

(ix) MERKMAL

30

(A) NAME/SCHLÜSSEL: Cytoplasmatische Domäne

(B) LAGE: 1-11

(ix) MERKMAL

35

(A) NAME/SCHLÜSSEL: Transmembran-Domäne

(B) LAGE: 12-30

(ix) MERKMAL

5

(A) NAME/SCHLÜSSEL: Stamm und katalytische Domäne

(B) LAGE: 31-325

(ix) MERKMAL

10

(A) NAME/SCHLÜSSEL: N-Glycosylierungsstelle

(B) LAGE: Asn 127

(ix) MERKMAL

15

(A) NAME/SCHLÜSSEL: Gesamtprotein

(B) LAGE: 1-325

20

```

      1 MKVFRRARWH RVALGLGGLA FCGTLLLYLA RCASEGETPS
    41 ASGAARPRAK AFLAVLVASA PRAVERRTAV RSTWLAPEER
    81 GGPEDVWARF AVGTGGLGSE ERRALELEQA QHGDLLLLPA
   121 LRDAYENLTA KVLAMLTWLD ERVDFEFVLK ADDDSFARLD
   161 ATLVDLRARE PARRRRLYWG FFSGRGRVKP GGRWREAAWQ
   201 LCDYYLPYAL GGGYVLSADL VHYLRLSREY LRAWHSEDVS
   241 LGTWLAPVDV QREHDPFRDT EYKSRGCNNQ YLVTHKQSPE
   281 DMLEKQQMLL HEGRLCKHEV QLRLSYVYDW SAPPSQCCQR
   321 KEGVP

```

30 Schema 3a

Dieses Schema zeigt die Aminosäuresequenz im Ein-Buchstabencode, welche durch Weglassen der cytoplasmatischen und transmembranalen Domänen aus der Sequenz Nr. 2a hervorgeht (Seq ID Nr. 3a).

35

```

    31 RCAAEPGDPR AMSGRSPPPP APARAAFLA VLVASAPRAA
    71 ERRSVIRSTW LARRGAPGDV WARFAVGTA LGAEERRALE
   111 REQARHGDDL LLPALRDAYE NLTAQVLAML AWLDEHVAFE
   151 FVLKADDDSF ARLDALLAEL RAREPARRRR LYWGFFSGRG
   191 RVKPGGRWRE AAWQLCDYYL PYALGGGYVL SADLVHYLRL
   231 SRDYLRAWHS EDVSLGAWLA PVDVQREHDP RFDTEYRSRG

```

40

271 CSNQYLVTHK QSLEDMLEKH ATLAREGRLC KREVQLRLSY
311 VYDWSAPPSQ CCQRREGIP

5

Schema 3b

Dieses Schema zeigt die Aminosäuresequenz im Ein-Buchstabencode, welche durch Weglassen der cytoplasmatischen und transmembranalen Domänen aus der Sequenz Nr. 2b hervorgeht (Seq ID Nr. 3b).

10

31 RCASEGETPS ASGAARPRAK AFLAVLVASA PRAVERRTAV
71 RSTWLAPEER GGPEDVWARF AVGTGGLGSE ERRALELEQA
111 QHGDLLLLPA LRDAYENLTA KVLAMLTWLD ERVDFEFVLK
151 ADDDSFARLD AILVDLRARE PARRRRLYWG FFSGRGRVKP
15 GGRWREAAWQ LCDYYLPYAL GGGYVLSADL VHYLRLSREY
231 LRAWHSEDVS LGTWLAPVDV QREHDPFRDT EYKSRGCNNQ
271 YLVTHKQSPE DMLEKQQMLL HEGRLCKHEV QLRLSYVYDW
311 SAPPSQCCQR KEGVP